

Temporal Relations Learning with a Bootstrapped Cross-document Classifier

Seyed Abolghasem Mirroshandel and Gholamreza Ghassem-Sani¹

Abstract. The ability to accurately classify temporal relation between events is an important task for a large number of natural language processing applications such as Question Answering (QA), Summarization, and Information Extraction. This paper presents a weakly-supervised machine learning approach for classification of temporal relation between events. In the first stage, the algorithm learns a general classifier from an annotated corpus. Then, it applies the hypothesis of "one type of temporal relation per discourse" and expands the scope of "discourse" from a single document to a cluster of topically-related documents. By combining the global information of such a cluster with local decisions of a general classifier, we propose a novel bootstrapping cross-document classifier to extract temporal relations between events. Our experiments show that without any additional annotated data, the accuracy of the proposed algorithm is at least 7% higher than that of the pattern based state of the art system.

1 INTRODUCTION

In recent years, much progress have been made in natural language processing (NLP). Combining statistical and symbolic methods has played a significant role in these advances. As a result tasks such as part-of-speech tagging, morphological analysis, parsing, and named entity recognition have been addressed with satisfactory results [1]. However, problems such as temporal information processing that requires deeper analysis of meaning are yet to be addressed.

Temporal information is encoded in the textual description of events. Lately, the increasing attention about the practical NLP applications such as question answering, information extraction, and summarization have resulted in a growing demand of temporal information processing [2]. In question answering, one may expect the system to answer questions such as "when an event occurred", or "what is the chronological order between some desired events". In text summarization, especially in multi-document type, knowing the order of events is a useful source for correctly merging related information.

For a given ordered pair of components (x_1, x_2) , where x_1 and x_2 are times and/or events, a temporal information processing system identifies the type of relation r_i that temporally links x_1 to x_2 . The relation type can be one of the 13 types proposed by James Allen [3]. For example, in the sentence "Travelers estimated (*e45*) that the California earthquake (*e46*) last month (*t47*) will result (*e10*) in a fourth-quarter (*t27*) pre-tax charge (*e49*) of less than \$10 million" (taken from document wsj_0144 of Timebank corpus [4]), there are

three relations between pairs (*e45, e10*), (*e46, t47*), and (*t47, t27*). The task is to automatically tag these pairs respectively with the *before*, *is_included*, and *before* relations.

With recent construction of the Timebank corpus [4], the efficiency of different machine learning methods can be now compared. The recent work with the Timebank corpus has disclosed that even a six-class classification of temporal relations is a complicated task, even for human annotators [1].

This paper presents a weakly-supervised machine learning approach for classification of temporal relation between events. In the first stage, the algorithm learns a general classifier from an annotated corpus. Then, it applies the hypothesis of "one type of temporal relation per discourse" and expands the scope of "discourse" from a single document to a cluster of topically-related documents. By combining the global information of such a cluster with local decisions of a general classifier, we propose a novel bootstrapping cross-document classifier to extract temporal relations between events. Our experiments show that without any additional annotated data, the accuracy of the proposed algorithm is at least 7% higher than that of the pattern based state of the art system [5].

The remainder of the paper is organized as follows: section 2 is about previous approaches to the temporal relation extraction. Section 3 explains our proposed method. Section 4 briefly presents the characteristic of the corpora that we have used. Section 5 demonstrates the evaluation of the proposed algorithm. Finally, section 6 includes the conclusions and some possible future works.

2 PREVIOUS WORKS

There are a lot of ongoing researches focused on temporal relation extraction. These efforts can be divided into three groups: 1) Pattern based; 2) Rule based, and 3) Anchor based. These groups are discussed in the following sub-sections.

2.1 Pattern based methods

This group of methods tries to extract some generic lexico-syntactic patterns for events co-occurrence. Extracting these patterns can be done manually or automatically.

2.1.1 Manual extraction of patterns

In Manual extraction of patterns, some patterns are defined manually. Perhaps the simplest pattern based method is the one that was developed using a knowledge resource called VerbOcean [6]. VerbOcean has a small number of manually selected generic patterns. The style of patterns is in the form of <Verb-X> and then

¹ Department of Computer Engineering, Sharif University of Technology, Tehran, Iran, emails: {mirroshandel@ce.sharif.edu, sani@sharif.edu}

<Verb-Y>. After creating these patterns, some of the semantic relations between events are obtained. A major weakness of this method, similar to other manual methods, is that the method tends to have a high recall but a low precision. One way to overcome this flaw is to create more specific patterns. However, it is clear that this solution would be both hard and time consuming.

The other way for resolving the low precision problem is using an additional component for pruning extracted relations. Many researches have tried to address this issue from a variety of point of views. For example, some used heuristics to compute the impact of precision [6, 7, 8]. Another approach was to incorporate a supervised classifier. There are some examples that have used this approach [9].

2.1.2 Automatic extraction of patterns

Automatic extraction methods use some machine learning techniques to extract patterns. They try to learn a classifier from annotated corpora, and attempt to improve classification accuracy by feature engineering. MaxEnt classifier is a good example of this group [1]. MaxEnt assigns one of six relations to each pair of events from an augmented Timebank corpus. This classifier uses a number of features that were hand-tagged in the corpus, including tense, aspect, modality, polarity, and the event class. In addition to these features, it also relies on two other features including pairwise agreement of tense and aspect.

In another approach, an event classifier is trained for inter-sentential events, and a corpus that contained sentences with at least two events, one of which triggered by a key time word (e.g., after, before, etc.), is built [10].

The state of the art of this group is very similar to the MaxEnt classifier. It relies on the features automatically extracted from raw text, and works 3% better than MaxEnt. This classifier tries to learn event attributes and event-event features in two consecutive stages. Event attributes are the same as that of MaxEnt, but event-event features are new and include part of speech, event-event syntactic properties, prepositional phrase, and temporal discourse [5]. Our focus is on this method and trying to improve it.

Another interesting method of this group employs a corpus of parallel temporal and causal relations, and explores a number of machine learning approaches to extract such relations [11]. Since existing corpora provide no parallel temporal and causal annotations, 1000 conjoined event pairs have been annotated [12]. The result of experiments shows that extracting causal and temporal relations mutually benefit from one another [11]. There are also some algorithms which utilize grammatical information in support vector machines classification using some novel convolution tree kernels. These methods have been shown satisfactory results in temporal relation classification [13, 14]. There is also an algorithm which uses cascaded finite-state grammars together with a machine learning component [15].

There are some other methods that have used machine learning techniques for semantic relation acquisition between events [16], which can also be applied to temporal relation extraction.

2.2 Rule based methods

The common idea behind rule based methods is to find some if-then rules for classifying temporal relations. In most existing work, the rules are determined manually based on Allen's interval algebra [3]. This algebra has some limitation that was pointed out by [17].

In a study, rules of temporal transitivity were applied to make

the size of training set ten times larger than original set. Next, the MaxEnt classifier was trained on this enlarged corpus. The accuracy test on this enlarged corpus was very encouraging [1].

Reasoning with determined rules is the other point of view of rules' usage. In [2], a rich set of rules (axioms) was created, and by using a first order logic based theorem prover, they tried to find a proof of a temporal relation by refutation.

2.3 Anchor based methods

Anchor based methods use information of argument fillers (i.e., anchors) of every event expression as a valuable clue for recognizing temporal relations between events. They rely on the distributional hypothesis [18] and by looking at a set of event expressions whose argument fillers have a similar distribution, try to recognize synonymous event expressions. Algorithms such as DIRT [19], TE/ASE [20], and that of Pekar's system [21] are some examples of such methods.

DIRT is an unsupervised method that is based on an extended version of Harris' distributional hypothesis [18]. Based on the original Harris' hypothesis, words that occur in the same contexts are usually similar. In DIRT, this hypothesis has been applied to paths in the dependency trees (i.e., If two paths tend to occur in similar contexts, the meanings of the paths tend to be similar.) [19].

TE/ASE is another unsupervised algorithm, which has two major phases. In the first step (called Anchor Set Extraction or ASE), the algorithm tries to extract similar anchors. In the second phase (called Template Extraction or TE), the system extracts templates from resulted anchor sets. After completing the second phase, some post-processing is applied to the extracted templates and some of improbable templates are omitted [20].

Because of complementary role of each group of methods, it seems that one can merge some of the methods discussed in previous sections to gain more accuracy. Recently some researches have focused on this issue. For example, pattern and rule based methods have been merged in [1, 22], and the new systems are more efficient than each of the base methods. In other examples, pattern based and anchor based methods have been combined [6, 16]. There is, however, one exception that combining pattern and anchor based methods has not shown satisfactory results, yet [7].

3 BOOTSTRAPPED CROSS-DOCUMENT CLASSIFIER

In this section, a novel method for extracting temporal relation between events is introduced, which have shown a significant improvement over previous works in terms of accuracy. Before describing the new method, our motivation is briefly explained in the next sub-section.

3.1 Motivation

In an ordinary corpus with heterogeneous documents, a particular verb can sometimes be an event trigger which may represent various event types. However, for a cluster of topically-related documents, the distribution may be much more convergent. This clue motivated us to apply the hypothesis of "one sense per discourse" [23] to the temporal relation classification and extend the scope of discourse from a single document to a cluster of topically related documents.

Inspired by the proposed work about one event trigger sense and one event argument role per cluster [24], we propose a similar hypothesis about one type of temporal relation per cluster. It means

that between two pair of similar (or synonym) events, only one type of temporal relation per cluster exists.

3.2 Feature engineering

In this research, two types of features are used: 1) Event attributes, and 2) Event-Event attributes.

3.2.1 Event attributes

Event attributes are the first type of features. For each event, there are five temporal attributes which are tagged in the Timebank corpus: 1) tense; 2) grammatical aspect; 3) modality; 4) polarity, and 5) event class. Tense and aspect define temporal location and event structure; thus, they are necessary in any method of temporal relation extraction. Modality and polarity specify non-occurring or hypothetical situations. The event class shows the type of event. The range of values for these attributes is based on [4], and it is shown in table 1. In addition to these five attributes, we also used part of speech tag of each event in our algorithm.

Table 1. The range of values for five temporal attributes

Attribute	Range of values
Tense	none, present, past, future
Aspect	none, prog, perfect, prog_perfect
Modality	none, to, should, would, could, can, might
Polarity	positive, negative
Event Class	report, aspectual, state, I_state, I_action, perception, occurrence

3.2.2 Event-Event attributes

These types of attributes are about two events occurring simultaneously, which includes event-event syntactic properties, prepositional phrase, and temporal discourse.

Event-Event syntactic properties: if both events were in the same sentence, we can use the parse tree of the sentence to learn useful properties. For utilizing these parse trees, we can use some tree kernels which are proposed in [13].

Prepositional phrase: preposition heads are often indicators of a temporal class, thus we can use a new feature that indicates if an event is part of a prepositional phrase.

Temporal discourse: it is an expected conclusion that the relationships between events are inversely related to the textual distance of the two events. It means that relationships become weaker as distance increases, and vice versa. Therefore, we decided to treat intra- and inter-sentential events differently. Thus we train two separate models, one for intra-sentential events and one for inter-sentential events.

3.3 Proposed algorithm

The proposed algorithm applies a novel usage of bootstrapping in temporal relations classification between events. Figure 1 shows the flowchart of the algorithm. It includes two main stages:

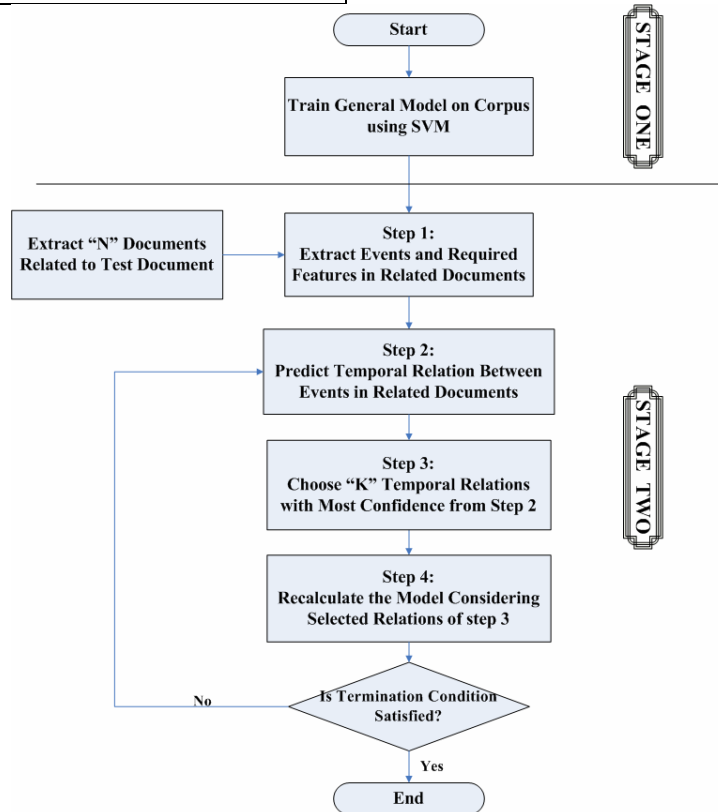


Figure 1. The flowchart of the proposed algorithm.

3.3.1 Stage one

In this stage, system uses specified features in a corpus to train a model for classification using Support Vector Machines (SVM). At the end of this stage, we will have a general model for temporal relation classification. Such models have been proposed before, however, they all had a problem of being too general, and thus not having any specific information about the particular domain under consideration. To overcome this problem, we propose an extra phase of training, which is discussed in the next section.

3.3.2 Stage two

In this stage, we try to retrain the general model produced in stage one with related information about the test domain. In order to achieve this goal, we try to obtain the top N (we tested various N, and N = 25 achieved best gains) related documents for each test document. We then should try to retrieve valuable information from these related documents. It can be done by a novel usage of bootstrapping according to the following steps:

Step 1: try to extract events and required features in related documents. The specified features of section 3.2 can be determined automatically by EVITA [29]. In spite of the fact that some of the events and/or features extracted by EVITA may be incorrect, our experimental results show that they can be very helpful.

Step 2: by the existing model, try to predict temporal relations between events in the related documents, and compute a normalized measure of confidence on the algorithm's prediction. SVM uses a distance measure, based on support vectors, to predict the proper class of test cases. Therefore, it is easy to compute confidence based on this distance measure (i.e., the closer a case is to the support vectors, the less it is confident). This distance can be computed based on following equation [25]:

$$d(X^T) = \sum_{i=1}^l y_i \alpha_i X_i X^T + b_0$$

where y_i is the class label of support vector X_i ; X^T is a test tuple; α_i and b_0 are numeric parameters that were determined automatically by SVM algorithm; and l is the number of support vectors.

Step 3: choose K (based on experiments, we selected K = 40) most confident temporal relations from step 2.

Step 4: consider selected temporal relations of step 3, and recalculate the model.

Step 5: repeat steps 2-4 until at least one termination condition is satisfied. The training will terminate if either no more unselected temporal relation remains, or validation accuracy no longer increases. Another simple condition for termination can be a predefined maximum number of iterations of these steps.

The fundamental idea of this stage is to obtain document-wide and cluster-wide statistics about the temporal relations between types of events, and then using this information to improve temporal relation identification and classification.

4 CHARACTERISTICS OF CORPORA

We used Timebank (v 1.2) and Opinion Corpus [1] in our experiments. Timebank has 183 newswire documents with 64077 words, and Opinion Corpus has 73 documents with 38709 words. These two datasets have been annotated based on TimeML [4]. There are 14 temporal relation types (Event-Event and Event-Time relations) in the TLink class of TimeML. Similar to [1, 2, 26], we used a normalized version of these 14 temporal relation types, which contains only the following 6 temporal relations:

- SIMULTANEOUS
- BEFORE
- ENDS
- IBEFORE
- BEGINS
- INCLUDES

For converting 14 relations into 6, the inverse relations were omitted, and IDENTITY and SIMULTANEOUS, as well as IS_INCLUDED and DURING were collapsed, respectively.

In our experiments, like some previous works [1, 5, 22], we merged two datasets to generate a single corpus called OTC. Table 2 shows the normalized TLink class distribution (only for Event-Event relations) over Timebank and OTC corpora.

As it is shown in table 2, relation "BEFORE" is the most frequent relation; thus it forms the majority class, and was used as the baseline of our experiments.

Table 2. The normalized TLink class distribution in TimeBank and OTC.

Relation Type	TimeBank	OTC
IBEFOR	63	131
BEGINS	77	160
ENDS	114	208
SIMULTANEOUS	1304	1528
INCLUDES	588	950
BEFORE	1335 (38.35 %)	3170 (51.57 %)
TOTAL	3481	6147

As it was discussed in section 3.3.2, for each text, we should retrieve 25 related texts. In our experiments, these related texts have been retrieved from English part of TDT5 multilingual news text corpus [27]. In total, TDT5 corpus consists of 407505 text documents in languages such as English (278109 documents), Mandarin Chinese (56486 documents), and modern standard Arabic (72910 documents). On the other hand, it also has 250 different topics. Unlike previous TDT corpora, TDT5 does not contain any broadcast news data; all sources are newswire. This corpus has applications like information detection, information extraction, language modeling, machine learning, machine translation, topic detection and tracking.

5 EXPERIMENTAL RESULTS

We have used LIBSVM [28] java source for the SVM classification, Evita [29] for event extraction, Stanford NLP package [30] for tokenization, sentence segmentation, part of speech tagging, and parsing. We have also used of INDRI retrieval system [31] to obtain related documents.

5.1 Experiments

We applied our algorithm to both Timebank and OTC corpora. Using the five-fold cross validation method, the results were

evaluated by measuring accuracy.

Figures 2 and 3 show the results of different algorithms on temporal relation classification between events over Timebank and OTC corpora, respectively. In these figures, the baseline is the majority class for event-event relations (i.e., the BEFORE relation) of the evaluated corpora. Mani's method is the most successful algorithm in the pattern based group of methods, which exclusively uses gold standard features of events [1]. Chambers' method is similar to Mani's method. Using extra features in a two step algorithm, it tries to improve Mani's method. Chambers is currently the state of the art method in pattern based approach. This algorithm uses some extra resource such as WordNet to achieve a higher accuracy [5].

Argument ancestor path distance (AAPD) kernel is an accurate convolution tree which only uses parse tree of event-event sentences for temporal relation classification [13]. AAPD polynomial is a composite kernel that combines Mani's method and AAPD kernel [13].

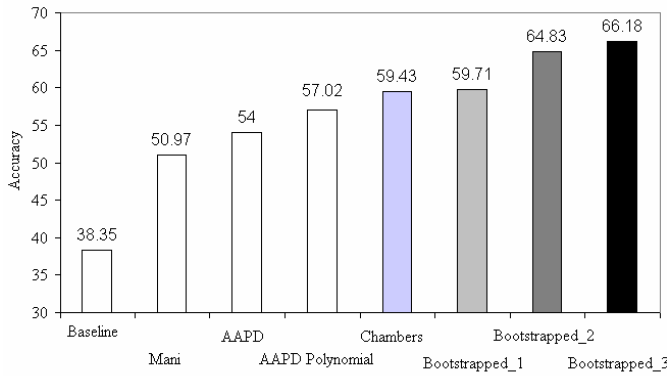


Figure 2. The accuracy of proposed methods on event-event temporal relation classification over Timebank corpus.

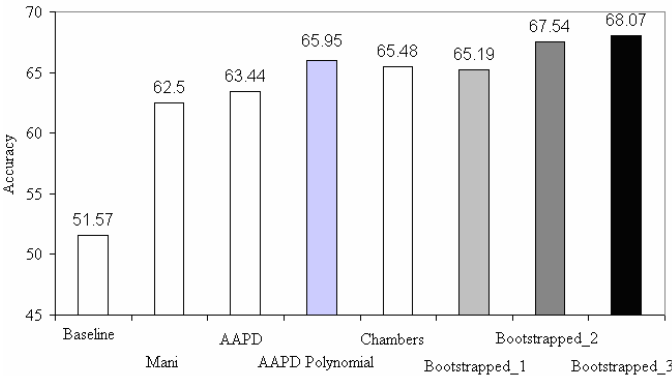


Figure 3. The accuracy of proposed methods on event-event temporal relation classification over OTC corpus.

Bootstrapped_1 method is our bootstrapped algorithm, which only uses event attributes. This method does not use event-event attributes. We can apply SVM to these simple features to construct our model. In bootstrapped_2 and bootstrapped_3, we utilized event-event attributes in addition to simple event attributes. For using syntactic property of event-event (the parse tree of the sentence containing two events), we have applied some effective tree kernels. CollinsDuffy tree kernel [32] and argument ancestor path distance (AAPD) tree kernel have been used in bootstrapped_2 and bootstrapped_3, respectively. CollinsDuffy tree kernel is a convolution tree kernel which effectively counts the

number of common subtrees between two comparing parse trees. AAPD tree kernel is analogous to CollinsDuffy kernel except that it uses different weighting functions in computing kernel value. In AAPD, the significance of subtrees can be measured using distance from event arguments of the temporal relation. In this weighting approach, the closer a node is to one of the arguments, the less it is decayed by the weighting function. In other words, the nodes which are placed near the relation arguments are more important than nodes far from arguments.

The kernels on simple event features and event-event features should be combined to form a better kernel. Two types of combination can be applied: linear or polynomial [33]. Because of superior results of polynomial composite kernels in temporal relation learning [13], we have used polynomial composition in our experiments.

As the table 3 shows, methods bootstrapped_2 and bootstrapped_3 show significant improvement over state of the art methods.

Unfortunately, there do not exist many researches in the pattern based temporal relation classification between events, and thus we could compare our work only with the Mani, Chambers, AAPD, and AAPD polynomial.

5.2 Analysis

Comparing existing methods; the result of applying our bootstrapped cross-document algorithm shows substantial improvement of temporal relation classification in term of accuracy, without using any extra annotated resources. Now, we can claim that our proposed method is the state of the art method in pattern based group. The result of experiments approves our hypothesis about one type of temporal relation per cluster.

Bootstrapping and retrieving related document have the following positive effects:

1) By knowing the relation between events, we can predict the relation types between analogous events, which may appear in related documents.

2) In related documents, the sentences count with similar pattern will increase, and the tree kernels can better extract information from parse trees. Indeed, SVM can be more informative.

3) Timebank and OTC corpora are rather small with few examples for each relation. This problem can affect the performance of relation learners. Some researches have tried to fix this problem by generating new relations via some temporal rules [1]. In our method, retrieving related documents and extraction of new temporal relations in these documents can increase the number of relations and improve the performance of temporal relation classification.

In the stage two of the algorithm, an important implementation issue is to choose an appropriate number (N) of related documents to retrieve. In order to select an appropriate value for N, we tried different numbers from 10 to 50 over Timebank. As it is shown in figure 4, the best accuracy was achieved in the case that the number of related documents was 25. Where number N increased over 25, the accuracy of the algorithm reduced. The reason was that as N gets larger, the unrelated events to the test events will considerably increase. In other words, since the process of retrieval and tagging of the related documents is performed automatically, as N gets larger, the probability of errors will also increase. As a result, unrelated patterns may be included in the learning process. That also causes a reduction in the accuracy of the results.

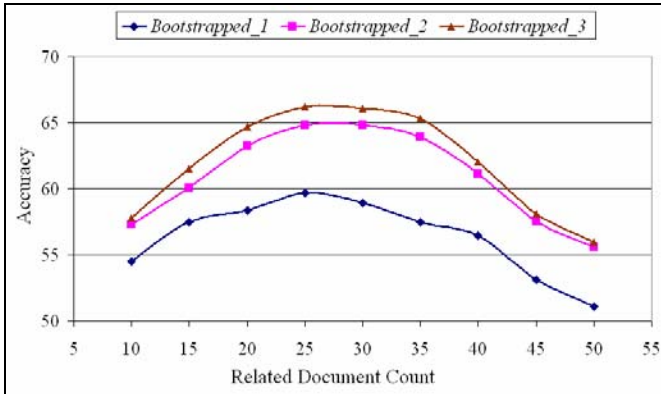


Figure 4. The effect of the number of retrieved documents on accuracy over Timebank corpus.

6 CONCLUSION AND FUTURE WORKS

In this paper, we have addressed the problem of temporal relation classification between events, which has been a topic of interest since early days of natural language processing. Here, we proposed a two-stage weakly-supervised approach for classification of temporal relations. In the first stage, the classifier is trained to learn temporal relations on the OTC corpus by SVM. Then, in the second stage, we apply the hypothesis "one type of temporal relation between events per discourse" via a cross-document bootstrapped classification algorithm. Based on this hypothesis, we could suggest incorporating global evidence from a cluster of related documents (new discourse) to refine local decisions from one document (old discourse). The results of applying the new method, without using any extra annotated data, showed a significant improvement over related works in the area of pattern based method (including the state of the art method) in terms of accuracy.

Currently, we are trying to find ways of further improving the accuracy of our system. It seems that using dependency structure of sentences or creating better kernels for SVM might be ways of doing so. Using WordNet synsets and discovering similar events is another possibility which may be useful for learning similar patterns between analogous event pairs from related documents.

REFERENCES

- [1] I. Mani, M. Verhagen, B. Wellner, and J. Pustejovsky, *Machine Learning of Temporal Relations*, In Proceedings of 44th Annual Meeting of the ACL, pp. 753-760. Australia, (2006).
- [2] M. Tatu and M. Srikanth, *Experiments with Reasoning for Temporal Relations between Event.*, In Proceeding of Coling-08, pp. 857-864. England, (2008).
- [3] J. Allen, 'Towards a General Theory of Action and Time', *Journal of Artificial Intelligence*, 23, 123-154 (1984).
- [4] Pustejovsky, J., Hanks, P., Sauri, R., See, A., Day, D., Ferro, L., Gaizauskas, R., Lazo, M., Setzer, A., and Sundheim, B., 'The TimeBank Corpus', *Journal of Corpus Linguistics*, 647-656 (2003).
- [5] N. Chambers, S. Wang, and D. Jurafsky, *Classifying Temporal Relations between Events*, In Proceedings of 45th Annual Meeting of the ACL, pp. 173-176. Czech Republic, (2007).
- [6] T. Chklovski and P. Pantel, *Global Path-based Refinement of Noisy Graphs Applied to Verb Semantics*, In Proceeding of IJCNLP-05, pp. 792-803, Korea, (2005).
- [7] K. Torisawa, *Acquiring Inference Rules with Temporal Constraints by Using Japanese Coordinated Sentences and Noun-Verb Co-Occurrences*, In Proceeding of HLT-NAACL'06, pp. 57-64, USA, (2006).
- [8] F.M. Zanaotto, M. Pennacchiotti, and M.T. Pazienza, *Discovering*

- Asymmetric Entailment Relations between Verbs Using Selectional Preferences*, In Proceedings of 44th Annual Meeting of the ACL, pp. 849-856. Australia, (2006).
- [9] T. Inui, K. Inui, and Y. Matsumoto, *What Kinds and Amounts of Causal Knowledge can be Acquired from Text by Using Connective Markers as Clues?*, In Proceedings of 6th International Conference on Discovery Science, pp. 180-193. Japan, (2003).
- [10] M. Lapata, and A. Lascarides, 'Learning Sentence-internal Temporal Relations', *Journal of Artificial Intelligence Research*, pp85-117, (2006)
- [11] S. Bethard and J.H. Martin, *Learning semantic links from a corpus of parallel temporal and causal relations*, In Annual Meeting of ACL-HLT, pp. 177-180, USA, (2008).
- [12] S. Bethard, *Finding Event, Temporal and Causal Structure in Text: A Machine Learning Approach*, Ph.D. thesis, University of Colorado at Boulder, (2007).
- [13] S.A. Mirroshandel, M. Khayyamian, and G.R. Ghassem-Sani, *Using Tree Kernels for Classifying Temporal Relations between Events*, In Proc. of the PACLIC23, pp. 355-364, Hong Kong, (2009).
- [14] S.A. Mirroshandel, G.R. Ghassem-Sani, and M. Khayyamian, *Event-Time Temporal Relation Classification Using Syntactic Tree Kernels*, In Proc. of the LTC'09 Conference, pp. 300-304, Poland, (2009).
- [15] B. Boguraev and R. K. Ando, *Timeml-compliant text analysis for temporal reasoning*. In Proc. of the 19th International Joint Conference on Artificial Intelligence, pages 997-1003, 2005.
- [16] S. Abe, K. Inui, and Y. Matsumoto, *Two-Phased Event Relation Acquisition Coupling the Relation-Oriented and Argument-Oriented Approaches*, In Proceeding of Coling-08, pp. 1-8. England, (2008).
- [17] M. Vilain, H. Kautz, and P. Van Beek, *Constraint Propagation Algorithms for Temporal Reasoning: A Revised Report*, In D. S. Weld and J. de Kleer (eds.), *Readings in Qualitative Reasoning about Physical Systems*, Morgan-Kaufman, pp. 373-381. (1989).
- [18] Z. Harris, *Mathematical Structure of Language*, John Wiley & Sons, New York. (1968).
- [19] D. Lin and P. Pantel, *Dirt: Discovery of Inference Rules from Text*, In Proceeding of the 7th ACM SIGKDD, pp. 323-328. (2001).
- [20] I. Szpektor, H. Tanev, I. Dagan, and B. Coppola, *Scaling Web-based Acquisition of Entailment Relations*, In Proceeding of EMNLP 2004, pp. 41-48. Spain, (2004).
- [21] V. Pekar, *Acquisition of Verb Entailment from Text*, In Proceeding of the Human Language Technology Conference of the NAACL 2006, Main Conference, pp. 49-56. USA, (2006).
- [22] N. Chambers and D. Jurafsky, *Jointly Combining Implicit Constraints Improves Temporal Ordering*, In Proceedings of EMNLP-08, pp. 698-706, USA, (2008).
- [23] D. Yarowsky, *Unsupervised Word Sense Disambiguation Rivaling Supervised Methods*, In Proceedings of 33th Annual Meeting of the ACL, pp. 189-196. USA, (1995).
- [24] H. Ji and R. Grishman, *Refining Event Extraction through Cross-Document Inference*, In Proceedings of 46th Annual Meeting of the ACL, pp. 254-262. USA, (2008).
- [25] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Second Edition. San Francisco: Morgan Kaufmann, (2006).
- [26] I. Mani, B. Wellner, M. Verhagen, and J. Pustejovsky, *Three Approaches to Learning TLINKs in TimeML*. Technical Report CS-07-268, Computer Science Department, Brandeis University, USA, (2007).
- [27] *TDT 2004: Annotation Manual*, Available at <http://www ldc.upenn.edu/Projects/TDT2004>.
- [28] C.C. Chang, and C.J., Lin, *LIBSVM: a library for support vector machines* (2001). Available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [29] R. Sauri, R. Knippen, M. Verhagen, and J. Pustejovsky, *Evita: A Robust Event Recognizer for QA Systems*, In Proceedings of the HLT05, pp. 700-707, Canada, (2005).
- [30] The Stanford Natural Language Processing Group, <http://nlp.stanford.edu/software/index.shtml>
- [31] T. Stohman, D. Metzler, H. Turtle, and W.B. Croft, *Indri: A Language-Model based Search Engine for Complex Queries (extended version)*. Technical Report IR-407, CIIR, Umas Amherst, USA, (2005).
- [32] M. Collins and N. Duffy, *Convolution Kernels for Natural Language*, In Advances in Neural Information Processing Systems 14, pp. 625-632, Canada, (2001).
- [33] M. Zhang, J. Zhang, J. Su, and G.D. Zhou, *A Composite Kernel to Extract Relations between Entities with Both Flat and Structured Features*, In Proceedings of ACL-44, pp. 825-832, Sydney, (2006).